图书情报工作
第 64 卷 第 15 期 2020 年 8 月

重大公共卫生事件的网络信息存档架构*
——新冠肺炎疫情的启示

■ 周文泓 苏依纹 吴琼 黄小宇 张晓宇 文丽君 贺谭涛
四川大学公共管理学院 成都 610064

摘 要:[目的/意义]针对重大公共卫生事件的网络信息展开存档研究,旨在从多维度探索网络信息存档的理论与方法。[方法/过程]基于代表性重大社会事件网络信息存档实践归纳参考框架,结合新冠肺炎疫情这一事件的具体信息情境,综合构建重大公共卫生事件的网络信息存档架构。[结果/结论]架构的基本内容为:一方面,阐释重大公共卫生事件网络信息的存档主体、存档信息对象、存档制度与技术保障、存档方案、网络档案库以及产品与服务组成的存档结果;另一方面,明确存档的总体要求:即表现为基于整体目标实施动态优化方案,实现事前全过程管控、全过程社会多元主体协作的基于国家视角的顶层设计与协同。

关键词:重大公共卫生事件 重大社会事件 新冠肺炎疫情 网络信息存档 存档架构
分类号:G270
DOI:10.13266/j.issn.0252-3116.2020.15.023

1 引言

基于以亿计的用户群体的网络成为世界各国与地区内各领域的活动发声、传播与践行平台,也是我国重大社会事件的信息即时发布、交流集成空间,汇聚了官方机构、主流媒体、利益相关社群、公众个人等多元视角信息。除了信息管理领域为探索各信息利益相关者应如何正面使用网络平台的问题而提出的应急情报、信息行为、后真相传播等命题外,通过网络被记录下的信息在事件过后的价值体现及处理方式同样值得深思。换言之,在我国已明确提出面向国家数字记忆实施互联网信息保存项目的背景下,这些庞杂的碎片化网络信息从历史叙事和信息资源产业的角度如何还原社会各界抗击疫情的多维过程,以及这些记录在传播过后如何作为国家记忆资源予以存续成为重要的研究议题,该议题可具象为基于新冠肺炎疫情的重大公共卫生事件的网络信息如何存档的问题。

实践层面,以传承国家信息资源与文化遗产为重要职责的档案馆、图书馆、博物馆等机构有在全国范围内发起围绕重大事件的资料征集传统。例如,新冠肺炎疫情的材料征集正在全国范围内展开。然而,在多数征集公告中,具有庞大规模的数字记录没有被明确提及,这些网络信息是否应存档、如何存档、由谁存档、存档哪些内容都缺乏整体规划。另一方面,国内外积累了重大社会事件存档的相关经验,图书馆、档案馆以及全球的网络信息存档联盟等机构参与其中,记录下的活动遍及各个领域。例如,欧洲难民危机[1]、加拿大联邦选举[2]、英国伦敦奥运会[3]等。这些实践呈现了基本的存档思路和方法,但均提出在范围设定、工具、法理与伦理风险、资源等方面存在局限,显示出重大社会事件的网络信息存档在司法、技术、社会、文化、管理等方面的复合挑战。

理论层面,对重大公共卫生事件网络信息存档的探讨一方面是以重大社会事件论证网络信息存档的必要性与价值[4];另一方面则是以实践案例,如伦敦奥运会,分析以事件为单位的网络信息存档挑战与策略[5]。同时,部分以网络信息或社交媒体信息存档为主题的研究提供了通用问题分析与方法指导基础,明确了网

* 本文系国家社会科学基金项目"基于多元数字技术的网络空间参与式归档研究"(项目编号:18CTQ037)研究成果之一。
作者简介:周文泓(ORCID:0000-0002-7496-5599),博士,副研究员,E-mail:zwhdcc@126.com;苏依纹(ORCID:0000-0002-1998-1887),硕士研究生;吴琼(ORCID:0000-0001-7304-5),本科生;黄小宇(ORCID:0000-0001-8173-3756),本科生;张晓宇(ORCID:0000-0002-8392-983X),本科生;文丽君(ORCID:0000-0001-6322-8780),本科生;贺谭涛(ORCID:0000-0001-9749-3638),本科生。
收稿日期:2020-03-13 修回日期:2020-05-29 本文起止页码:184-193 本文责任编辑:杜杏叶

络信息存档的基本要素:明晰存档必要性、合规依据和主体权责制度[6-9];主体机制,即存档的利益相关者及其分工框架[10];流程,涵盖设计归档范围/收集策略、捕获、存储和组织、处置、质量保证、利用等环节的具体步骤[11-14]等;技术,落实存档流程所需应用的工具和平台如捕获工具、检索算法[15-16]等。然而,重大公共卫生事件网络信息存档同为以"主体"生成者展开的网站或网络账户存档在信息对象和方法上存在不同。此外,随着社交媒体信息占有更大比例,原有的以网站,如政府平台为主的研究能提供的借鉴有限,社交媒体类的研究所提出的方案较少落实到具体的"事由"层来说明重大公共卫生事件如何存档,也缺乏公共卫生领域的实例参考。

因之,本研究将总结代表性的重大社会事件网络信息存档实践,形成供公共卫生事件参考的网络信息存档框架;同时,结合新冠肺炎疫情这一事件及其信息情境探讨我国重大公共卫生事件网络信息存档结构。

2 基于代表性案例的重大社会事件网络信息存档参考框架

经搜索引擎检索、档案馆与图书馆网站及 IIPC 网站(IIPC 为全球由有着收集、保存和提供利用互联网信息专长的图书馆、档案馆、博物馆等形成的协作网络,即国际互联网保存联盟)调查可发现,美国、加拿大、英国、德国、丹麦、乌克兰、中国、新加坡、马来西亚、澳大利亚等均有代表性的网络信息存档项目[1-3,17-29],也建成了大大小小的重大社会事件数字记忆资源库,涵盖政治、文化体育、经济、社会等各领域,覆盖从官方组织、主流机构到社会群体与个人形成的各类网络信息。由此,在成效与挑战中为重大社会事件网络信息存档提供参考框架。

为获得总结性参考框架,本文的依据与思路如下:①对实践项目的背景、内容、成果的线上调研结果进行文本转换,形成项目的具体描述。②结合档案学客体全宗将信息之间的来源联系明确为同一事件过程中信息的有机关联,以此认识重大社会事件的信息对象的前提[30-31]。依据网络信息存档的基本要素如主体、对象、方法、流程(包括捕获、整合、保存、发布等),对项目具体内容进行分类提取,结合各实践项目可供提取的内容,经由归纳和调整,明确了存档主体、网络信息对象、存档方案、存档结果、存档保障这五大要素及其分要素。③对各实践项目在不同要素下的内容进行整合,从而明确每一要素及其分要素在重大社会事件情境中的具体表现,并归纳出框架要点。

2.1 代表性案例介绍

当前存档成果可用且存档行动提供一定介绍的 20 余个重大社会事件的网络信息存档项目显示,从存档主体的规模来看形成以下几类代表性实践:

2.1.1 国际协作主导存档类

此类实践强调通过国际协作获得方法、资源与工具的共享,形成全球性力量对跨国性重大事件进行网络信息存档。以国际互联网保存联盟内容开发组(以下简称 IIPC CDG)为例[24],来自 45 个国家与地区的图书馆、档案馆等机构展开协作旨在建立有关国际重大社会事件的跨国的、多语言的、多观点的公开访问网络档案汇集。IIPC CDG 目前已展开的协作收藏主题包括奥运会、残奥会、欧洲难民危机、气候变化等具有国际影响的事件,其目标捕获的网络信息形式多样,包括网站、文章、新闻报道、网络日志以及 Facebook、Twitter 等社交媒体平台;最终馆藏涵盖了多个国家、多个语种网络内容。例如 2016 年里约奥运会项目最终馆藏涵盖 125 个国家,记录了 34 种不同的语言。该网络档案收藏项目以公开提名、公众决策的方式进行收集和保存,IIPC 成员机构与公众在公开网络会议讨论后均可通过公开提名贡献自己的建议网站列表。项目整合过程包括质量管控与组织描述,即通过爬网工具报告检查、人工审查两种形式的检查筛选,并对元数据进行详尽的著录。最终,重大社会事件网页馆藏在公共平台发布后进入后续研究程序,如 IIPC CDG 提取奥运会网络档案数据中的性别数据进行奥运会参与者性别分布研究。此过程中,IIPC CDG 主要使用第三方合作开发工具与平台服务来保障其存档程序,其 2010-2014 年奥运会收藏项目的收集工具为北德克萨斯大学开发的 URL 提名工具,网站的发布与保存依托于 Internet Archive 的 Archive-it 平台服务。

2.1.2 国家级主导存档类

此类实践基于国家视角,围绕产生国家网络舆情、具有一定研究价值的重大社会事件活动进行的网络信息捕获、保存与开放利用活动,强调由具有法定存缴义务的国家记忆机构主导开展。以大英图书馆的 UK web archive 项目[3]和英国国家档案馆的 UK government web archive 项目[26]中面向 2012 伦敦奥运会的存档实践为例,英国构建关于奥运会的网络信息资源仓储原因有二:①从本国视角出发,记录 2012 伦敦奥运会的盛况,构建本国民众的奥运会国家记忆;②在世界范围

内,为全世界人民留存关于奥运会的记忆资源。两个项目所捕获的网络信息涵盖历届奥运会的举办状况、比赛报道、典型任务、事例报道等,为公众提供全方位的网络信息资源。具体行动上,UK web archive 中奥运会相关网络信息存档则是先确定了捕获的主题及范围,进行捕获后将其以确定主题直接进行整合和发布。在这过程中,依靠于本身项目在主体权限、资源获取、资源访问等方面的制度保障,项目均能顺利开展与实施。在技术实现方面,两者目前均使用开源工具与企业技术服务来捕获奥运会相关内容,对于转发等互动性内容的存档技术仍待开发。

2.1.3 地区或机构级主导存档类

此类实践围绕对于地区、机构产生重要影响的事件进行网络信息存档活动。存档主体主要为相应的记忆机构或部门。此类实践基于地区视角、机构职能活动视角、用户视角来选择存档事件,收集平台广泛。以加拿大魁北克省国家图书馆和档案馆网络档案项目(BAnQ)为例[27],BAnQ 将网站存档视作其完成魁北克出版遗产材料保存使命的一部分,希望利用网络档案支持魁北克省历史与当下社会情境的留存。存档对象为反映魁北克省特定历史时期社会生活的网络信息,其中包括围绕重大事件产生的网络内容,如生态事故、社会选举、纪念活动等事件。BanQ 主要采取手动捕获方式进行事件型网络内容收集,且按照事件强度、网络信息传播范围采取不同捕获频率:政治选举等定期事件采取"周捕获",生态事故等突发事件在发生初期内则选择"日捕获"策略;捕获后的信息经过价值、版权、质量等多方面的严格筛选与整合后方可存入云端存储器,并最终以列表形式发布于 BanQ 门户网站上,同时提供主题与 URL 搜索功能。技术与制度保障层面,项目的存档与重映工具(Heritrix、OpenWayback)均为国际范围内认可度、使用度最高的开源软件,魁北克省国家图书馆和档案馆作为加拿大法定存缴单位,拥有收集权限,但公布网络存档行为仍需获得信息制造方许可。

2.1.4 社群级主导存档类

此类实践选择对社群发展起重要作用的事件进行存档,主要收集社群成员围绕事件产生的一系列网络记录内容,存档活动一般在形式上有统一标准要求,但内容上则呈现丰富、包容的社群与个人特色。以 GitHub 平台上的"2020 新冠肺炎疫情个人故事"项目为例,其存档主体为 Github 上的个人用户群体。在目标维度上该实践可以看作是中国普通公众在疫情期间的个人生活记录备份。该实践选择豆瓣平台网络日志为存档对象,大部分记录来自湖北武汉,反映了抗疫期间武汉普通市民的真实生活,构建成珍贵的社群网络记忆。项目采取网友共建的行动策略,网友按照统一格式、程序抓取豆瓣日记页面中的日记正文与图片部分进行保存,合并保存时需对文件进行基本描述,著录项包括文章转载授权、作者、来源与发布时间,其存档内容公开发布于 GitHub 平台[25]。

2.2 参考框架发现与解析

基于对现有重大社会事件网络信息存档项目的归纳分析,形成了相对通用的行动框架(见图 1),具体如下:

2.2.1 框架内容:存档要素及其关联

从实践来看,相对完备的重大社会事件网络信息存档框架包括如下要素及其作用:

(1)存档主体,是重大社会事件信息的利益相关者,主要包括发起者与协同者。发起者也可视为存档的主导者或主要行动主体,贯穿倡议、启动、统筹、设计与实施等全流程的活动。实践中发起者主要为两类,一类是传统的记忆机构如档案馆、图书馆、博物馆等,它们是具有专业能力与配套资源的信息保管者。另一类为较具社会记忆建构意识与能力的社会第三方,例如信息生成者、传播者、利用者以及可通过参与存档受益的提供资源和技术支持的社会第三方机构、群体或个人。作为存档主体,发起者与协同者基于职责、目标以及能力共同发现重大社会事件及其网络信息,深化对信息形成情况、价值以及特征的认识,依此形成网络信息存档需求与行动的基本认知,为存档形成组织机制,明确存档目标与方案。

(2)网络信息对象,即重大社会事件在网络平台的相关记录。它是存档信息的来源,也是存档的作用对象,依据事件性质对群体和深度进行记录。它影响着信息形成的基本情况如数量、形式、内容等,体现了不同的价值与特征,这对判定存档主体的存档必要性、程度、所需资源和具体方案都产生影响。其中,更为重要的是通过相对全面的调查帮助梳理和完善重大社会事件的主要脉络和维度,为从背景层理解并认识信息和整体关联提供基础。

(3)存档保障,用以全程指导与落实存档方案的实施。制度在广义上包含用以指导重大社会事件网络信息存档的规则集合,帮助形成存档方案。重大社会事件网络信息往往具有跨背景、多信息生产者的特征,因而信息边界与权责难以简单划分, 全方位存档指导

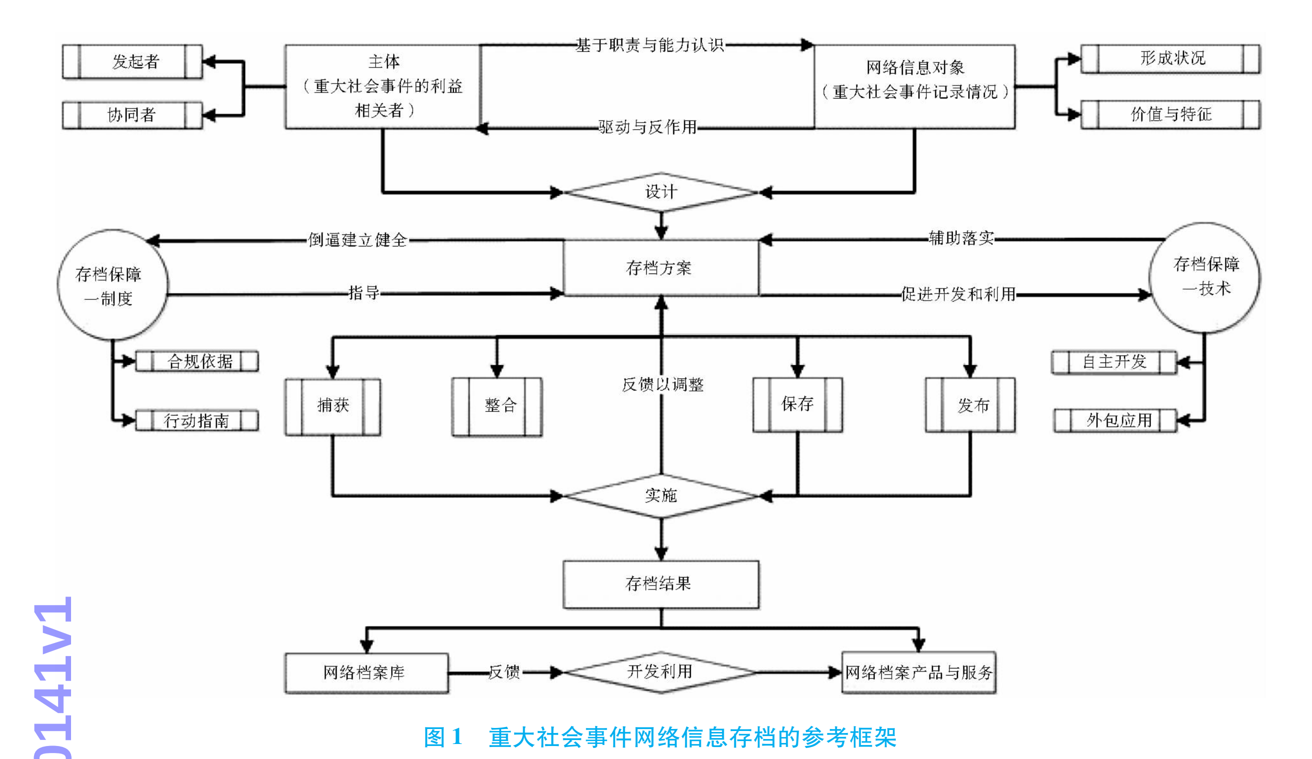


图1　重大社会事件网络信息存档的参考框架

不可缺失。制度的作用主要体现在两个方面，首先是为存档从必要性和风险规避提供合规依据，如信息所有权、利用权、隐私、知识产权等。然后是为存档提供行动指导。从理论基础、方法、原则、规范等层面落实存档行动的整体模式、流程与活动细节。例如，基本所有存档项目设置删除制度，确保用户若发现项目网络档案存在侵犯隐私内容可联系负责人删除。

另一个存档保障是技术，它是落实存档方案的关键辅助。网络信息存档从方案到落地实施亟需技术的配套使用，从捕获到发布乃至开发利用均需技术支持。存档主体既可根据需求自主开发，也可向专业领域寻求“外包”支持或改造共享开放的技术。

（4）存档方案，用以落实捕获、整合、保存、发布的过程要素总设置。存档方案的设计是一种过程性活动，例如确认存档范围、捕获频率与方式、信息保存格式、质量标准、发布条件等。需要说明的是，整合是将网络信息有序化的过程，包括组织、描述、鉴定等具体环节，且网络信息的特征是开放，因而保持合规的“线上可见”特质也是存档重要内容。同时，存档方案将随着实施效果及反馈情况予以动态调整，并倒逼制度的建立健全与技术的优化完善。

（5）存档结果。一是网络档案库，即存档方案实施的直接成果，是留存重大社会事件记忆与证据的信息资源集成。它的内容与形式直接反映存档方案的实施成效。二是网络档案产品与服务。除了向社会提供重大社会事件的网络信息原材料，存档主体可逐渐通过深层次开发，形成“信息密度”更高的产品与服务。

2.2.2　框架要点

为连接好各要素形成完备的重大社会事件网络信息存档框架，且为了充分应用框架，存档基本策略体现的基本理念和方法须贯穿其中。

（1）参与和协作是构建组织机制的基本要义。其一，有力的主导核心是得以面向重大社会事件实现网络信息存档的重要主体力量，这往往需要几大条件：存档意识和“志愿”，即认识到重大社会事件网络信息的价值并愿意落实于行动，例如英国为应对潜在的“数字黑洞”于2004年成立网络档案馆，旨在收集、访问和保存来自英国域的具有学术和文化重要性的网络资源[3]；存档能力，指具有从专业方案出发做好存档统筹与执行的能力，并兼具管理、技术、人文等各方面能力；存档资源，即能为存档配置相应的人才、资金、基础设施等支持；存档职责，往往有法定或社会认同应留存社会记忆资源的机构、群体或个人可实现倡议和行动统筹。其二，重大社会事件网络信息往往是跨平台、跨群体、跨形式甚至跨国而形成，传播路径糅合于复杂的社交与内容网络中，信息权责所要覆盖的利益相关者广泛，加之存档工作的复合专业度需求，均要求社会多元主体的参与。例如，IIPC的存档内容提名机制[24]是CDG领衔和社会广泛参与提名二者的平衡，既借助社会力量寻获广泛信息来源，又确保了面向社会记忆建

构的整体统筹。再如,将工具应用从自主开发拓展向更多开源或是部分有偿工具的支持,在一定程度上解放了存档机构的技术压力。

(2)深化存档结果"有限完整"的认知。存档而成的档案库同重大社会事件网络信息相比本就只是其中一部分,部分信息价值有限,有的甚至是噪音数据,抑或是不适合纳入具有一定"公共性"的资源整体中。其次,技术与资源限制难以做到全部存档,如杜克大学档案收集政策[19]中提到的,由于服务的结构,某些社交媒体平台很难或无法收集,且由于社交媒体服务不断变化,不能保证能够持续收集所有资料。因而存档阶段性结果并不等同于存档方案的最终目标,存档方案的每个具体环节往往要精细化设置,例如捕获要依据存档范围中的价值或是难易度等确定哪些信息的优先级更高,整合和发布也是如此,同批捕获的信息也可能由于各类原因整合进度与发布时间不一致。总而言之,存档结果要容许"完整性"上的局限。例如,英国政府网络档案馆(UKGWA)项目[26]自身明确指出,由于技术限制,现阶段可存档的社交媒体信息不包括 Facebook 平台,以及所有平台中交互性的内容(如转发、评论等),所以现阶段对某一事件的刻画也是不尽完善的,而将来是否继续开展、如何开展面向重大社会事件的网络信息存档将会考虑政府的需求、资源的现实状况等多方面因素。

(3)明确存档是非线性的持续过程。由于重大社会事件网络信息的记录广度与深度对存档提出的从资源到具体行动上的挑战涉及社会、文化、技术、管理、司法等多个方面,它的存档难以实现线性化的一蹴而就,而是要在动态中调整完善,事前、事中、事后均要有存档行动。事前存档主要在于要形成重大社会事件网络信息的常规机制,为形成针对性的方案提供通用的框架基础,如魁北克国家图书馆和档案馆网络档案(BAnQ)[27]、澳大利亚潘多拉档案库[28]、英国网络档案馆(UKWA)[3]等项目都列出了其较为完备的常规网络信息存档框架,来作为具体方案的指导性文件;事中的存档要确保及时性,这是针对网络信息较易失存、失真的特性而言,且事中信息的存档需面向网络环境中事件不断复杂化的信息关联;事后存档的对象是那些需要长时捕获、跟踪捕获的信息,并且此类信息后续的处理和发布等同样耗费时间。此外,存档方案的事前设计和存档行动在事中实施的过程中也会呈现不足,即便是信息源的发现也需事后完善。美国国会图书馆关于 2010 年巴西总统选举的网络信息存档就从 2010 年 9 月一直持续到 2011 年 1 月,自大选前就开始捕获相关的预热消息,在大选结束后也持续收集民众的反映以及后续的报道,从而更加完整地展现实际状况[29]。

(4)制度与技术完善并举,不可偏废。重大社会事件网络信息存档的工作量与复杂性在实践中充分体现,当前并未有实践成果成为重大社会事件网络信息存档范围的完全对照。首先,存档从框架到针对性方案的设计都有赖于制度支持,这需要从理论上分析重大社会事件网络信息从社会、文化、管理、司法等维度体现的存档难点与痛点,并从方法层面明确应对要点,以此映射于制度层。例如,存档中涉及的隐私和信息所有权问题需要立足法学视角,而庞杂的信息关联分析则需依托档案学领域将"背景"与"有机关联"的内涵进一步扩充。这些均要迁移至制度层进行设定才可确保存档方案的科学性。其次,面对大量信息繁杂的信息存档活动,社会广泛的人工参与也不能完全满足工作量需求,大众参与带来的信息质量审核和价值判定同样有不小的工作量。例如,当前 GitHub 上的新冠肺炎疫情众包项目成效有限[25],一部分原因就在于仅信息真假判定要耗费大量时间。因而,技术全方位应用于存档已是不可或缺的策略。例如,从捕获到发布,IIPC 建议可用的工具就有 84 个[24]。

(5)存档单元以"事件"为中心,把握"事件"特质。相较于多数存档实践以形成主体为中心展开存档,重大社会事件的网络信息存档从认识信息形成情况到存档行动都更注重"事件"维度。这要求在信息对象的调查中不限于信息生成者这一视角,而是从事件本身出发来关联事件相关主体、场景、活动、演进脉络等要素。换言之,本质是要记录事件的哪些方面和哪些进展,不同记录主体只是其中的一个审视与存档维度。例如,部分网络信息存档尽管在整合结果上体现出不同事件,但由于存档时事件视角有限,所存档信息零散缺乏关联进而导致事件还原度不足。例如,Webarchiv 项目的最终成果仅限按字序排列的公开 url 列表,缺乏更深入组织,无法形象展示事件的更多面貌[20]。

3 我国重大公共卫生事件的网络信息情境——以新冠肺炎疫情为例

公共卫生事件由于关乎人的生存需求而拥有极大的社会影响力,其广泛的关注基数借助于当下丰富的网络平台不断扩张,形成规模庞杂的信息传播现象,以 2019 年末开始逐渐得到关注的新冠肺炎疫情为例,体

现为以下几方面基本情况：

（1）重大公共卫生事件如新冠肺炎疫情的网络信息主要形成于两类网络平台。一是自建网站，主要是各类机构主体如党政机关、新闻媒体、其他有一定规模的社会组织等。各层级、各系统的党政机关网站与各方新闻平台开辟了疫情信息专栏，内容主要包括每日疫情通报、防控新闻动态、通知公告发布、疫情知识普及、医务人员风采等，这类信息由官方机构权威发布，具有较强的参考价值[32]。央视网、新华社、人民网等也均有专题报道甚至直播[33-34]。二是由网络服务提供商支持的第三方平台如微博、微信、今日头条、抖音、快手、豆瓣、知乎。基于社交媒体本身的特点，广泛的用户群体形成了数量庞大且多种多样的疫情相关网络信息。同样以新冠肺炎疫情为例，主体上说，既有各级各类组织机构的官方账户，也有广大与疫情息息相关的遍布各领域的个人账户，从医护人员、志愿者到普通民众账户均包含其中。内容方面来说，由这些广泛的主体形成的网络信息反映了疫情的各个方面。例如，新冠肺炎疫情中应用最广的社交媒体平台——新浪微博显示，关于新冠肺炎疫情形成的话题带动了以百亿计的阅读量和千万级的讨论量：形成“新型肺炎超话”“肺炎患者求助超话”等超级话题；“隔离日记”、“抗击新型肺炎第一线”、“武汉加油！”等常规化微博话题；每日均有热搜话题同新冠肺炎疫情相关；在“发现页”专门开辟的“武汉日记”专区来记录疫情中人们生活各方面情况。

（2）原始记录性的信息比重高。新冠肺炎疫情的到来无形中倒逼各领域提升无纸化运行程度，网络空间的跨时空和多元化属性得到充分体现，从政府网站到社交媒体再到线上办公工具都发挥着不同的作用。这也使得新冠肺炎疫情的网络信息不仅在形式上趋异，且原始记录内容占据重要组成。例如，医护人员通过抖音、快手、微博等社交媒体以视频、图片、文本等形式记录下一手的抗疫相关行动。再如，远程办公已成为此次疫情决策与具体行动安排的重要工作方式，且钉钉、腾讯会议等平台提供了直播回放等功能，依托网络空间的涉及疫情的公务活动可实现全过程记录。这些重要的“业务”信息不是由其他形式转换或从其他设施迁移而来，而是直接在网络平台中记录而生成。

（3）主体维度的“共同生成者”特性强化。从新冠肺炎疫情的社会影响力来看，共同生成者从宏观上看是以亿计的网民，无论是生成者规模还是持续时间，较以往都极为罕见。新冠肺炎疫情从受感染者、救护人员、决策者、基层行动者、各领域与公众协同者等都在疫情中贡献不同方面的力量，在网络空间因业务需要、信息公开、公共参与等因素围绕疫情形成不同方面的记录，也会在同一活动中形成各自视角的信息。例如，微博新冠肺炎疫情超话是可能受感染求得救援的重要渠道之一，有着大量的病状描述、求助以及后续的反馈信息，1 000 多条求助贴子以及每个贴子下从几十到成千上万的转发或评论共同形成了关于求助和救援的部分记录[35]。

（4）内容映射了新冠肺炎疫情影响下的社会百态，超出了疫情本身的范畴。这是指作为中国举国协同应对重大危机的社会事件，从中央到地方、从决策层到基层，从卫健到各领域、从官方到社会公众的全体行动，由此形成无处不在的新冠肺炎疫情相关网络信息。这些信息不仅关乎疫情在卫健这一职能下所涉及的具体业务与活动，且包含所有面向疫情防控与应对的社会活动。因而，网络信息如要从整体来看待，是围绕各个活动及其涉及的主体相互协同交错所关联而成的记录体系，从内容上看涵盖不同时空下多样化的疫情相关主题。

（5）价值与挑战兼具。从价值来看，网络空间作为新冠肺炎疫情抗击的重要阵地，所形成的数量大、范围广、内容多样的信息具有高度的证据、记忆、资源与资产价值，从积累疾控经验启示、卫健系统优化、应急管理体系建设、展示和推进我国治理能力现代化、讲述中国故事、促进数字中国记忆构建等方面均可发挥作用，且体系如此庞大，仅新浪微博就有超过百个热搜话题与亿级讨论量的信息如何在跨平台、跨主体、跨形式、跨内容的前提下实现整体存档也为档案社会化全景实践提供试验场。

从挑战来看，数量庞大的网络信息应从何处获得配套的存档资源；哪些主体应当与能够存档这些信息，信息权利与义务如何识别和划定；信息噪音、低价值信息、重复信息或违法违规的信息发布账户如何识别；信息围绕疫情活动与实践主体产生的关联如何充分揭示；存档方案如何设计、存档范围如何设定、信息价值如何判定、存档行动优先级如何设置；有何法理与伦理风险，如何应对。这些挑战都需要首先解决一个基本问题，重大公共安全事件网络信息存档架构是怎样的？即存档规则体系已有哪些建构，有哪些缺失，以及跨平台、跨主体、跨形式、跨内容的网络信息如何有相对统一的存档框架？这决定了相应存档行动需关注哪些要素，且可以提供一个解决上述问题的基本讨论范畴。

4 我国重大公共卫生事件的网络信息存档架构

重大公共卫生事件的网络信息在我国应如何存档？一方面，我国网络信息存档有一定政策与实践基础。国家图书馆作为国际互联网保存联盟的成员已有长久的探索经验，领衔面向国家数字记忆的互联网信息社会化保存项目。国家图书馆于4月22日启动了涵盖数字信息的战“疫”记忆库建设[36]，在国家档案局2019年所发布的《政府网站网页归档指南》指导下[37]，各地方档案馆与网站主办单位正在展开存档行动。另外，重大社会事件网络信息存档在世界范围内亦有多个国家与地区展开的实践，尽管在现有公开材料里有关公共卫生事件的信息成果有限，但加拿大[2]、英国[3]等对政治、体育、社会灾难类事件的网络信息存档在重大社会事件的范畴内都可供参考。同时，新冠肺炎疫情提供了具体的信息情境。因而，立足我国国情，结合前文的重大社会事件网络信息存档框架，为着重解决当前缺乏基本方案的局限，应首先确认公共卫生事件的网络信息存档架构中的要素和总体要求。

4.1 存档要素的基本设定

4.1.1 存档主体方面

首先，当前以守护信息资产为职能的档案馆、图书馆、博物馆应发挥领衔作用。当前各地的档案馆、图书馆、博物馆正向社会征集新冠肺炎疫情的各类材料，鼓励社会不同程度的参与，这也应当拓展向网络信息对象中。国家图书馆发起的战“疫”记忆库项目[37]也明确提出将网络信息纳入捕获范畴，并提出各级记忆机构和社会各方力量的协作。因而，在目前分散的行动之下可见新冠肺炎疫情网络信息存档可成为图书馆、档案馆、博物馆建立全国联网体系共建共享存档资源的作用空间。其次，从政府网站到各类第三方平台，各网络服务提供方，尤其是主要通过API共享信息的社交媒体服务提供商，应成为存档的重要辅助者和资源方，例如应急管理部门、卫生健康部门等官方网站，以及发挥重要传播与发声作用的新闻媒体与社交媒体平台，均应通过社会倡议和官方对接的方式最大程度纳入协作体系中。再次，社会第三方，尤其是学术团体、技术组织、可提供存档基础设施的企业等可成为重要补充。最后，社会公众应成为持续的“供应”力量，这既在于社会公众作为网络信息的记录者和被记录者应有一定权利决策信息留存与删除，也在于社会公众协同能发挥单一方难以企及的群体智慧、资源和“工作力”。例如，IIPC CDG发起的协作收集项目在各个阶段多通过公开网络会议收集公众存档建议与提名存档网站[24]。

4.1.2 网络信息对象方面

需做好盘查与梳理工作。展开全方位调查，对网络信息形成平台、账户、信息数量、信息增长情况等方面的总体把握；新冠肺炎疫情同信息相互对照，明确有哪些方面的信息对应着新冠肺炎疫情相关的活动，从主体、平台、内容等角度做好关联梳理，从而形成面向新冠肺炎疫情这一重大社会事件的信息映射脉络。例如，从信息背后的社会活动来看，解析可从卫健与应急领域展开，辅助外围领域中哪些领域更为接近哪些相对远离都应得到描摹，各领域的行动从决策、业务到职能辅助须展开同疫情核心的远近度评估。同时，限于疫情没有确定的时间节点，存档行动往往在初步梳理完事件脉络后才能展开，为避免事件发展过程中信息累积过程造成存档困难，存档主体应在事中追踪碎片化主题并抓取信息，确保背景描述的完整性，从而为疫情后期或结束后的整理工作带来便利。

4.1.3 存档保障方面

制度方面，当前我国已有《网络安全法》《全国人民代表大会常务委员会关于加强网络信息保护的决定》《网络信息内容生态治理规定》《网络资源归档格式标准GB/T 33994－2017信息和文献WARC文件格式》等从法律到标准，从法规到部门规章的制度基础，但关于信息确权、存档合规性、长期保存、信息泄露防范等有待完善。同时，更加具体的指南规范目前尚需档案机构、图书馆、博物馆等主导方提供相对统一的依据，存档术语、方法、范围、流程与主体协同机制等皆是需在制度层明确的要点。

技术方面可从三个方面寻获支持，一是将开源共享的技术进行适应性完善，从IIPC分享的全流程配套多元工具可根据其特性选择合适的使用[24]，GitHub等技术开源平台同样也有可参考使用的工具，亦可通过类似主题的项目找到攻克共同问题的协同者[25]；二是Internet Archive向世界各国与地区提供的有偿存档工具如archive it也可供部分参与主体使用[38]；三是自主开发或是联合社会其他力量共同开发。技术不仅要用于存档全流程中各活动的需要，比如信息的自动识别、真实性鉴定、捕获、组织、描述等，而且要面向多元主体协同的需要建成可互联、互信、互通的网络化平台，从而得以将社会化参与的成果以整体方式进行呈现。

4.1.4 存档方案方面

存档方案整体上要确保从信息调研、捕获、整合、保存、发布各环节的完整设定与标准化衔接，并配套可执行的具体标准、依据，如存档范围、捕获频率与方式、信息保存格式、质量标准、发布条件等，这在已有的网络信息存档中可获一定参考。同时以新冠疫情为代表的公共卫生事件着重关注：一是存档范围的设定，面对记录范围如此之广、深与复杂的新冠肺炎疫情，全部存档并不现实，那么就要求平衡信息背后的多元形成主体、平台、形式、内容，行动者需做好细致的分类、价值判定以及行动优先级设定，且可对已有项目进行调查或可协商共享已捕获的信息；二是存档不仅要明确哪些网络信息对象应纳入档案范畴，且要厘清哪些信息限于质量、价值或是保护利益相关者权益如隐私等方面的原因应排除在范围之外并获得相应处置，疫情期间涉及的诸多个人隐私信息应得到有效保护；三是存档方案须细致以确保技术需求清单可实现完整对照制作；四是存档方案须留有空间，疫情何时停止或是如何发展难以确定，具有随机性的长线存档应确保方案调整的可能性，且随着存档经验的积累，相应优化也是必要的。

4.1.5 存档成果方面

新冠肺炎疫情网络档案库是存档的直接结果，可将其定位为：全面、开放、可映射。首先，要全面呈现新冠肺炎疫情全过程中各层级、各领域通过各网络平台展示的行动全貌；其次，开放是网络信息存档的要义，在技术条件可及和合规保障利益相关者权益的情况下尽量以开放数据标准提供在线利用；最后，档案库要实现对新冠肺炎疫情的映射，即不仅要提供所存档的信息，所呈现信息应是多维分类的结果，从而可从不同脉络展示新冠肺炎疫情。同时，在网络档案产品与服务的开发方面，新冠肺炎疫情网络档案库要建设相应的关联化平台，它不应只是资源保管与提供基本利用的渠道，更应具有“纪念馆”和“实验室”的功能，除了向利用者提供加工成果，也协同利用者针对这些信息资产深度挖掘为可服务于社会不同方面需求的产品，实现社会价值最大化。

4.2 存档的总体要求：国家视角下强化顶层设计与协同

新冠肺炎疫情这样的公共卫生事件的网络信息于国家而言具有重大价值，它的存档复杂性也有赖于从国家战略层提供资源、组织架构、行动指南等多方面多角度的协同规划。然而，我国除了社会层面有分散的存档行动如华中科技大学主导的战“疫”数字博物馆[39]、技术专长的社会个人在 GitHub 上开展的抗疫期中国群众普通生活记录项目外[25]，国家层面提出的更大规模、更多视角、内容更为丰富的项目则是国家图书馆在 2020 年 4 月 22 日发起的联合全国各级各类图书馆的中国战“疫”记忆库建设项目[36]，该项目明确指出疫情防控中运用的新技术和社会管理形态的变化，如数字化生活、网上办公、线上学习等被设置在捕获范畴之中。然而，相比于 IIPC 形成的可在线检索的获得网站访问许可 5000 条存档网页资源[24]，我国的行动表现出时间滞后和社会参与有限的不足。同时，存档各要素及其分要素要跨情境、跨主体、跨时空、跨对象落实为整体的行动同样需要有效的协调机制。由此，国家层面布局顶层设计与展开行动的时空协同就显示出它的必要性，体现为：

（1）基于整体性目标实施动态优化方案。鉴于新冠肺炎疫情网络信息涉及范围的广度、记录深度以及信息关联复杂度，存档对不同参与主体而言既要对具体目标进行细化分解，亦要有面向国家数字记忆构建的整体目标实现战略与方向指引。无论是分目标还是整体大目标，存档方案绝非即时达成匹配性方案，要设定适应于不同阶段不同存档条件与要求的，面向不同参与者的过程性方案，循序渐进地落实存档行动。这意味着方案的完备度是立足具体时空而言，并非绝对，且允许受制于客观条件与能力限制的“妥协性”不足。例如，目前难以将不同格式的数据完全以整体的方式对照于原记录同时写入. md 文件中，要允许以要素完整为先的“分散式”捕获。

（2）存档须实现事前性的全过程管控。网络信息形成速度极快且传播网络复杂，为避免传播之后信息生态的无序化堆积和随机失真失存，应尽早展开综合性部署规划，在疫情未完全解决，信息将持续增长的情况下需同步制定存档框架与方案。对于此次疫情的网络信息存档相对滞后的情况，应及时总结经验，纳入国家记忆机构如图书馆、档案馆业务活动升级与制度建设的范畴中，为公共卫生事件乃至重大社会事件的网络信息存档构建前置框架与通用预案。此次新冠肺炎疫情网络信息在记录主体、数量、内容多样性、平台多元性等方面更加复杂，其发生与发展较为随机。在已滞后的情况下行动要最大程度紧凑展开，可试点先行，以点成面形成整体布局和全景行动。

（3）存档全过程有赖社会多元主体协作。网络信息的复杂性意味着应发挥多方力量实现体系化与规模

化存档,其中包括了以守护国家信息资产为职责的档案馆、图书馆、博物馆;可提供和核心记录的信息形成主体如卫健与应急管理部门、医院及其前线救护人员、社区、志愿者等;提供技术和资源的第三方如媒体、高校学术团体、企业、非营利组织等;参与互联网信息保存的且设立了疫情专题的网络平台服务商;在我国优势体制下最大化鼓励公众参与,将新冠肺炎疫情这一万众参与的事件落于社会共建式、还原式的存档实践中。

5 结语

总而言之,面向重大社会事件的网络信息是极具价值的证据、记忆、资源与资产,受国家治理能力现代化与国家数字记忆构建而驱动系统存档。本文结合网络信息调研结果与代表性重大社会事件存档提供的参考框架,从主体、对象、方案、保障、结果等要素出发,基于新冠肺炎疫情情境设计公共卫生事件的网络信息存档架构。同时,网络信息存档架构尚需由要素与总体要求的基础结构扩充为更系统的框架,有待形成更具体的方案与实践验证,需要后续升华探讨网络信息存档在事件维度下的理论以及方法要义。

参考文献:

[1] Royal Library of Denmark. Netarkivet. dk[EB/OL]. [2020 - 02 - 20]. http://netarkivet. dk/in-english/.
[2] LAC. GCWA[EB/OL]. [2020 - 02 - 20]. http://webarchive. bac-lac. gc. ca/#c.
[3] THE BRITISH LIBRARY. UK Web archive[EB/OL]. [2020 - 02 - 20]. https://www. webarchive. org. uk/en/ukwa/index.
[4] 马宁宁,曲云鹏,谢天. 欧洲主要网络资源采集项目研究与启示[J]. 图书情报工作,2013,57(12):10 - 15.
[5] 仇壮丽,许冬玲. 归档网络信息选择策略的影响因素研究[J]. 档案学研究,2011(3):63 - 66.
[6] 周毅. 论网络存档权及其生成[J]. 中国图书馆学报,2011,37(1):102 - 108.
[7] 赵俊玲,杜国芳. 著作权法对网络信息资源保存的影响分析[J]. 现代情报,2005(5):72 - 74.
[8] VLASSENROOT E, CHAMBERS S, DI PRETORPO E, et al. Web archives as a data resource for digital scholars[J]. International journal of digital humanities,2019,1(1):85 - 111.
[9] 周文泓,陈怡,张玉洁,等,王璠. 英国国家档案馆网络归档的案例分析及其启示[J]. 档案管理,2018(4):4 - 7,74.
[10] 赵俊玲,卢振波. 网络信息保存的责任体系分析[J]. 大学图书馆学报,2006(2):94 - 97,88.
[11] 吴硕娜,黄新荣. Web 归档生命周期模型的发展研究[J]. 数字图书馆论坛, 2018(10):41 - 45.
[12] 陈清文. 网络信息资源长期保存的采集策略与方法[J]. 情报探索,2006(12):47 - 48.
[13] 王文玲,曲云鹏. 网络存档数据质量保证策略理论框架研究[J]. 知识管理论坛,2018,3(2):106 - 115.
[14] 马宁宁,曲云鹏. 中外网络资源采集信息服务方式研究与建议[J]. 图书情报工作,2014,58(10):85 - 89,116.
[15] 唐光前. 网络资源的自动获取识别归档技术研究[J]. 现代图书情报技术,2003(5):57 - 61.
[16] GOSSEN G, RISSE T, DEMIDOVA E. Towards extracting event-centric collections from web archives[J]. International journal on digital libraries,2018:1 - 15.
[17] TAMIMEN LIBRARY OF NEW YORK UNIVERSITY. Special collections[EB/OL]. [2020 - 02 - 20]. https://specialcollections. library. nyu. edu/search/? f% 5Brepository_sim% 5D% 5B% 5D = tamwag.
[18] UNIVERSITY OF TEXAS AT AUSTIN LIBRARY. Web archive [EB/OL]. [2020 - 02 - 20]. http://legacy. lib. utexas. edu/hrdi/.
[19] DUKE UNIVERSITY ARCHIVES. Uawebarchive[EB/OL]. [2020 - 02 - 20]. https://library. duke. edu/rubenstein/findingaids/uawebarchive/#collectionoverview.
[20] NATIONAL LIBRARY OF THE CZECH REPUBLIC. Webarchiv [EB/OL]. [2020 - 02 - 20]. https://www. webarchiv. cz/en/.
[21] NATIONAL ELECTRONIC ARCHIVES OF UKRAINE. Collections [EB/OL]. [2020 - 02 - 20]. https://tsdea. archives. gov. ua/en/.
[22] DEUTSCHE NATIONALBIBLIOTHEK. Web archive [EB/OL]. [2020 - 02 - 20]. https://www. dnb. de/EN/Professionell/Sammeln/Sammlung _ Websites/sammlung _ websites _ node. html # doc272684bodyText1.
[23] NATIONAL LIBRARY OF NEW ZEALAND. New Zealand web archive[EB/OL]. [2020 - 02 - 20]. https://natlib. govt. nz/collections/a-z/new-zealand-web-archive? search% 5Bpath% 5D = items&search% 5Btext% 5D = web + archive.
[24] INTERNATIONAL INTERNET PRESERVATION CONSORTIUM. CDG collections [EB/OL]. [2020 - 02 - 20]. https://netpreserveblog. wordpress. com/? s = Helena + Byrne.
[25] JIAYILIUJIAYI. 2020nCov individual archives[EB/OL]. [2020 - 02 - 20]. https://github. com/jiayiliujiayi/2020nCov_individual_archives.
[26] NATIONAL ARCHIVES. Webarchive [EB/OL]. [2020 - 02 - 20]. http://www. nationalarchives. gov. uk/webarchive/.
[27] BANQ. Services [EB/OL]. [2020 - 02 - 20]. https://www. banq. qc. ca/services/bibliotheque_nationale/depot_legal/collecte_sites/.
[28] NATIONAL LIBRARY OF AUSTRALIA. Pandora overview[EB/OL]. [2020 - 02 - 20]. http://pandora. nla. gov. au/overview. html.
[29] LIBRARY OF CONGRESS. Web - archiving[EB/OL]. [2020 - 02 - 20]. https://www. loc. gov/programs/web-archiving/about-

this-program/.

[30] 冯惠玲,何嘉荪.全宗理论的实质——全宗理论新探之二[J].档案学通讯,1988(5):8-11

[31] 蔡娜.重大事件档案管理机制研究[J].档案学通讯,2012(3):65-67.

[32] 中华人民共和国国家卫生健康委.疫情通报[EB/OL].[2020-02-20]. http://www.nhc.gov.cn/xcs/xxgzbd/gzbd_index.shtml.

[33] 新华媒体创意工场.疫情数据[EB/OL].[2020-02-20]. http://fms.news.cn/swf/2020_sjxw/3_12_worldYQ/index.html? v=0.8829852148310229.

[34] 人民网.疫情报道[EB/OL].[2020-02-20]. http://society.people.com.cn/GB/369130/431577/index.html.

[35] 新浪微博.#新冠肺炎疫情#[EB/OL].[2020-02-20]. https://s.weibo.com/weibo? q=%23%E6%96%B0%E5%86%A0%E8%82%BA%E7%82%8E%E7%96%AB%E6%83%85%23.

[36] 国家图书馆.国家记忆 共同守护国家图书馆关于启动中国战"疫"记忆库项目暨征集抗击新冠肺炎疫情主题资源的公告[EB/OL].[2020-02-20]. http://www.nlc.cn/dsb_zx/zyjyk/.

[37] DA/T 80-2019,政府网站网页归档指南[S].北京:国家档案局,2019.

[38] INTERNET ARCHIVE. Archive-it[EB/OL].[2020-02-20]. https://www.archive-it.org/.

[39] 华中科技大学出版社.武汉战"疫"数字博物馆[EB/OL].[2020-02-20]. https://lofterhuati.lofter.com/tag/%E6%AD%A6%E6%B1%89%E6%88%98%E2%80%9C%E7%96%AB%E2%80%9D%E6%95%B0%E5%AD%97%E5%8D%9A%E7%89%A9%E9%A6%86.

作者贡献说明:

周文泓:提出研究问题、思路,撰写论文,修改论文;
苏依纹:查找、分析案例,撰写论文;
吴琼:查找案例,修改论文;
黄小宇:查找案例,修改论文;
张晓宇:查找案例,修改论文;
文利君:查找文献;
贺谭涛:查找文献。

Internet-based Information Archiving Architecture of Major Public Health Events ——Enlightenment from Covid-19

Zhou Wenhong Su Yiwen Wu Qiong Huang Xiaoyu Zhang Xiaoyu Wen Lijun He Tantao
School of Public Administration, Sichuan University, Chengdu 610064

Abstract: [**Purpose/significance**] Research on archiving Internet-based information from the perspective of major public health events aims to explore theory and method of web archiving in pertinence dimension. [**Method/process**] With summarizing the reference frame based on the representative archiving practices and analysis of the information status of Internet-based information from the perspective of COVID-19 event, a framework for archiving Internet-based information of major public health events was comprehensively proposed. [**Result/conclusion**] Therefore, on the one hand, explain the archiving subject, information object, system, technology, scheme, Internet archive, Internet archive products and services of Internet-based information of major public health events. On the other hand, clarify the general requirements of archiving: as shown in implementation of the dynamic optimization scheme based on the overall objective, prior control of the whole process, cooperation of multiple social subject, construction of the Chinese model under the open framework.

Keywords: major public health event information of major social event novel coronavirus pneumonia Internet-based information archiving archiving architecture